CRT-2001-12

Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

ECONOMIES OF SCALE IN EMPTY FREIGHT CAR DISTRIBUTION IN SCHEDULE RAILWAYS

by

Martin Joborn
Teodor Gabriel Crainic
Michel Gendreau
Kaj Holmberg
Jan T. Lundgren

Mars 2001

102-02934/

PUBLICATION
CRT-2001-12

*ECONOMIES OF SCALE IN EMPTY FREIGHT CAR DISTRIBUTION IN SCHEDULE RAILWAYS**

by

Martin Joborn[1]
Teodor Gabriel Crainic[2]
Michel Gendreau[3]
Kaj Holmberg[1]
Jan T. Lundgren[1]

* This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR).

1 Department of Mathematics, Linköping University, Linköping, Sweden
2 Département de management et technologie, Université du Québec à Montréal, et Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7
3 Département d'informatique et de recherche opérationnelle et Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7

Dépôt légal – Bibliothèque nationale du Québec
Bibliothèque nationale du Canada, 2001

Abstract

In this paper we consider empty freight car distribution in a scheduled railway system. We analyze the cost structure for the repositioning of empty cars, and conclude that the distribution cost shows an economy of scale behavior. In addition to the cost proportional to the number of cars sent from origin to destination, there is a cost related to car handling operations at yards which depends on the number of car groups that are handled. Thus, if we can find a transportation pattern in which fewer but larger groups of cars are built, the total distribution cost will decrease.

The objective of the paper is to propose an optimization model that explicitly takes this economy of scale effect into account. We use a time-expanded network to describe the possible car movements in time and space, and show how this network can be transformed into a network with fixed costs on links representing movements of cars with identical origin and destination terminals. The resulting optimization model is a capacitated network design model, where each capacity constraint limits the flow on several arcs. We describe a tabu heuristic for solving the model and present computational results.

Key words: Empty freight car distribution, Time-dependent capacitated network design, Tabu Search, Scheduled railways

Résumé

L'objet de cet article est la distribution des wagons vides dans un système ferroviaire avec horaires. L'analyse des coûts révèle l'existence de certaines économies d'échelle associées au nombre de groupes de wagons manipulés dans les gares de triage. Les coûts de distribution pourraient ainsi être diminués en augmentant la taille des groupes de wagons et en réduisant leur nombre.

Nous proposons un modèle d'optimisation pour le problème de distribution des wagons vides qui traite explicitement ces économies d'échelle. Un réseau espace-temps permet de représenter les mouvements des wagons, et des trains, et de formuler un modèle de synthèse de réseau multi-produits avec capacités restreignant les flux de plusieurs arcs à la fois. Nous proposons une méta-heuristique avec tabous pour résoudre ce problème et présentons des résultats expérimentaux.

Mots-clés : Distribution des wagons vides, synthèse des réseaux espace-temps avec capacités, recherches avec tabous, services ferroviaires avec horaires

1 Introduction

In freight transportation systems, there are often spatial imbalances in the supply and demand of empty transportation equipment, such as trailers or freight cars. This is a result of regional imbalances regarding the flows of goods. To compensate for these imbalances, empty transportation equipment has to be repositioned from areas with an excess of transportation equipment to areas with a deficit. This paper considers empty freight car distribution in a scheduled railway system. The empty freight car distribution process includes the problem of planning and performing the movements of empty freight cars with the objective of minimizing cost while satisfying the demand and supply for empty cars. Empty transports represent a substantial part of the total workload in a railway system, and an efficient empty distribution process is very important for reducing both operating costs and capital costs. Since there are many types of freight cars, and every car type cannot be used to transport all types of goods, empty freight cars may have to be transported to a terminal even if there is already an excess of empty cars at that terminal. This fact emphasizes the importance of an efficient car distribution system.

Some of the costs for empty transports in a railway system are proportional to the number of freight cars moved and directly related to transportation distances. Other costs are caused by the freight car handling at yards, for example costs for classification. These costs constitute a large portion of the total cost for empty distribution, especially in smaller systems. The costs for car handling operations often depend on the number of *car groups* that are handled at the yards, where a car group is one or more cars that are held together during the operations and therefore can be treated as one handling unit. If fewer but larger groups of cars could be handled together at each yard, the shunting and classification processes would not only be simplified but it would also decrease the associated per unit car cost. Thus, the cost structure shows an economy of scale behavior. Most models for empty freight car distribution, such as those reviewed in [1] and [4], do not explicitly take this economy of scale into account. As a result, these models may yield freight car distributions that involve movements of many small car groups.

The purpose of this paper is to formulate and solve an optimization model for empty freight car distribution that explicitly considers the economies of scale in the distribution process. The railway system we consider is a scheduled railway, in which all trains are running according to a timetable. The timetable defines arrival and departure times, capacities of trains, which days of the week the trains are run, etc. We assume that there exists information about available transportation capacities in the trains and also reservation possibilities for freight cars in the trains. The timetable together with the reservation possibilities certifies that the empty transports can be performed as planned, making the implementation phase of the empty car distribution deterministic. The modeling of economy of scale is made by introducing fixed costs on certain arcs in the network, in addition to the arc costs proportional to flow. The proposed model is dynamic in the

sense that a time-dimension is included in the model and it considers multiple time periods. The model will generate distribution plans with coordinated car movements so that the number of car groups and number of car handling operations on freight cars are reduced. Thus, the total distribution cost will decrease. The optimization model is intended for operational planning, but it can also be used at the tactical and strategic planning levels. The motivation for the work comes from SJ Cargo Group, a division of Swedish Railways, and the planning situation at this company is the background to the formulation of the proposed model. Some assumptions about the problem require that the information systems at the railway company are rather well developed, which is the case at Swedish Railways. Although we have the Swedish situation in mind, we are convinced that the problem scenario is similar in many other railway companies, and that the modeling approach therefore is applicable also in a general setting, especially at railway companies that run trains according to detailed pre-specified schedules.

To summarize, the contributions of the paper are: the analysis of economies of scale in empty freight car distribution, the optimization model for the empty freight car distribution problem, and the metaheuristic approach to optimize the model.

The remainder of the paper is organized as follows. In Section 2 the empty freight car distribution problem is described, and the cost structure of the process is discussed in Section 3. This is followed in Section 4 by the modeling assumptions and a description of how we introduce fixed costs in the model to capture the economies of scale. A mathematical model is also presented. Section 5 describes a tabu heuristic for solving the model and numerical results are presented. Finally, the paper is concluded in Section 7.

2 Problem description

In this section we describe the empty freight car distribution problem in a scheduled railway. The general planning process for empty freight car distribution at a railway company is described, and we discuss the economies of scale effect in the planning. Related literature is reviewed.

2.1 The planning process

In the planning process at a railway company we can identify three planning levels. At the strategic level the size of the resources is determined, at the tactical level the usage of the given resources is planned, and at the operational level the actual operations of the resources are determined. In this paper, we focus on the operational planning level.

The flow distribution of loaded and empty cars in the railway network constitutes a *transportation pattern*. It includes the specification of the origin and the destination of all freight car movements, assignment of freight cars to trains and temporal aspects of the transport. The operational costs depend strongly on the transportation pattern. Moreover, to analyze the costs of a transportation pattern, given the available system resources, one has to study the correlation between the tactical and operational planning levels.

The most important part of the tactical planning in a scheduled railway is to plan and construct the timetable. The timetable defines days of operation, routes, stops, arrival and departure times, and capacities of the trains. It also defines which blocks are built at yards and which train connections are possible. The handling times at classification yards are taken into account in the timetable. As a consequence, the timetable defines all feasible ways to perform the transport of cars from an origin to a destination. The transportation time on each route between two terminals, including handling times at terminals and classification yards, is also known. The timetable is usually revised a few times a year.

We define a train by its route, origin, destination, intermediary stops and physical path, as well as the day of operation and departure and arrival times at all stations. This means that if the timetable specifies that one train shall run a specific route each day, we consider this as multiple trains, one for each day. Thus, each train is run only once.

The general freight transportation pattern is determined simultaneously with the schedule during tactical planning. One of the most important decisions in operations planning is then to update, instantiate, and implement the transportation pattern given the actual demand, given the timetable. The capacities of the trains are determined at the tactical planning level, and this capacity has to be shared between loaded and empty freight cars. As soon as the demand for loaded transports is known to the planner, the corresponding loaded freight cars are booked into the trains. Most of the demand for transports is known at least one day before the loaded transport is to be performed. The remaining capacity of the train after loaded cars are booked can be used for empty cars. When planning the empty car distribution, the loaded transports cannot be altered, so the remaining capacities of the trains for empty freight car transports are fixed.

The loaded transports generate a demand for (and a future supply of) empty freight cars. The need for empty freight cars is known at least one day before demand for cars have to be fulfilled, since the demanded loaded transports are known one day in advance. In smaller railway systems, such as Swedish Railways, the span between the time which a demand for empty cars is known and the time when it has to be fulfilled is longer than most transportation times, which means that the demand for empty cars can be considered as known in the empty freight car distribution process. The objective of the operational planning of the empty freight car distribution is to satisfy as many demands

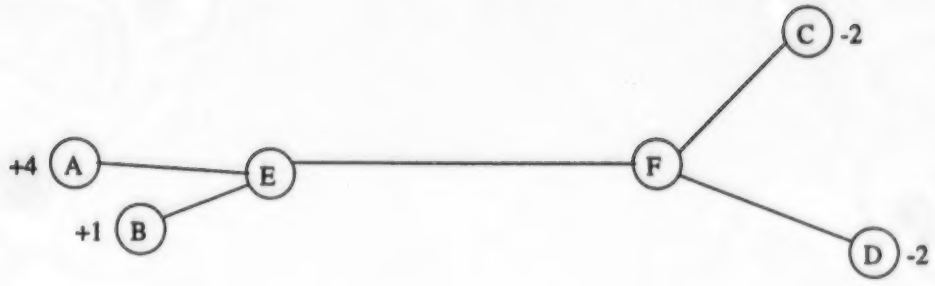

Figure 1: Illustration of transportation trade-offs

for empty cars as possible, while minimizing the distribution costs.

Compared to loaded transports, the planning of empty transports gives a much larger freedom for the railway company in how to implement the empty movements. A transportation order of loaded freight cars has a fixed origin and a fixed destination and is often scheduled for a specific day, which means that there are very few alternatives for implementing the order and the movement. When repositioning empty freight cars, there are no origin-destination specific demands that have to be fulfilled, and no specific time schedule for each empty transport. The only requirement is to satisfy the total demand, given the supply for each car type at each terminal.

The freedom of how to move empty freight cars can be used to coordinate the empty transports in order to obtain a cost efficient transportation pattern that explicitly accounts for economies of scale.

The coordination idea is illustrated in Figure 1, using a small railway network with only four terminals (A,B,C,D) and two classification yards (E,F). At terminals A and B there is a supply of empty cars of four and one units, respectively, and at each of terminals C and D there is a demand for empty cars of two units. The distances between the terminals and yards are proportional to the lengths of the arcs in the figure. The total transportation distance will be minimized if two cars are transported from A to C, one car from A to D, and one car from B to D. An alternative solution is to send two cars from A to C and two cars from A to D. The total distance transported will then increase, but this transportation pattern will simplify the work (car handling) at terminals B, E and F. If the increase in the transportation distance is not too large, many railway practitioners would prefer the second solution. In this paper we want to develop an optimization model for empty freight car distribution that reflects the relative gains associated to this type of consolidation/coordination of empty freight car movements.

2.2 Literature review

Several models for empty freight car distribution have appeared in the literature. Ssee [4] for a review of such models. Optimization models for train routing and scheduling in general are surveyed in [1].

Most of the models for empty distribution do not explicitly take economies of scale into account. Instead, it is implicitly assumed that the transports already at a tactical planning level are organized to utilize some large scale advantages. This means that the transports are scheduled so that each origin-destination specific movement is implemented in a cost-efficient way. The routing of the freight cars on the trains is not explicitly included in the models. For example, Powell and Carvalho [13] determine the number of flatcars to be sent from one terminal to another by formulating a logistics queueing network [12]. Various aspects are considered in the decision process, but it is assumed that the cost of sending cars is proportional to the number of cars. Holmberg, Joborn, and Lundgren [8] make a more detailed description of the transportation possibilities in a railway system. They include assignment of cars to trains in the model, but they do not include consideration of large scale effects consolidation/coordination in the car distribution.

Haghani [7] formulates a model for tactical planning of train routing, train makeup, and empty freight car distribution simultaneously. The estimated flow of empty cars is considered in the tactical planning and influences the train routing and makeup decisions. In this model the economies of scale and empty distribution are taken into account at the tactical level.

In [14], Turnquist explicitly takes the economies of scale in empty freight car distribution into consideration. Turnquist formulates the empty freight car distribution problem as an uncapacitated network design problem. The design arcs correspond to the origin-destination (OD) pairs utilized in the distribution. The objective is to find a distribution that combines minimizing the transportation distance and minimizing the number of OD-pairs in which empty cars are sent.

Comparing the goal of the optimization modeling in this paper with these works, we find that while Haghani approaches the problem at the tactical planning level, we have the operational planning stage in mind, considering the trains as stipulated. Further, while Turnquist does not include any transportation capacities, we include capacities on the trains. Train capacities are a central planning issue at many railway companies (see [8], for example). The train capacities make an important difference to the network design approach used by Turnquist. While Turnquist models each possible OD-flow with an arc, we recognize that several trains may be used to perform a transport between an origin and a destination (and conversely, transports between many OD-pairs may be transported in the same train), making it impossible to model a train capacity with

simple flow restriction on an arc representing an OD-flow.

3 The cost structure

In order to design an efficient transportation pattern, it is necessary to analyze the cost structure in the empty freight car distribution process, and to identify the relevant costs which directly or indirectly are associated with a given pattern. We are interested in the costs that we can influence by changing the transportation pattern, i.e., costs that depend on the flow and routes of freight cars. Costs that do not depend on how we design the transportation pattern (for example maintenance costs, certain personnel costs and capital cost) are not considered. We can also distinguish between costs that depend on the decisions at the tactical planning level, and the costs that depend on the operational planning.

In this section we describe and analyze the most important costs for the empty freight car distribution. We first describe the costs corresponding to trains, and costs for moving cars that are directly related to transportation distance. The latter type of cost is included in most models for empty freight car distribution. Further, we describe the car handling costs related to classification at yards.

Some costs in the railway system are not relevant to the construction of the transportation pattern, since they cannot affect the design of the transportation pattern. For example, we consider the holding cost to be independent of the terminal of which cars are held. As a result, the total holding cost in the railway system is not alterable in the operational planning. We therefore do not include any holding costs in the optimization model. However, a holding cost may easily be included in the model if desired.

3.1 Train cost

An important share of the total costs for implementing a transportation pattern is associated with the trains. Some train costs are fixed and do not depend on what is transported in the train, for example engine and driver driver costs. Such costs are related to the decision to run a train or not. Other costs, such as energy costs, depend on the load of the train and the transportation pattern. We denote these variable train costs as transportation costs, which are discussed below.

The train costs can be modeled as a fixed cost that is positive if the train is run, otherwise it is zero. In a scheduled railway, the largest portion of the trains is determined in the tactical planning stage, and only a very minor part of the total train cost can be

influenced by the transportation pattern that is determined at the operational planning stage.

Since the optimization model is aimed at operational planning, we ignore these fixed train costs in the model, assuming that they are determined at the tactical planning stage.

3.2 Transportation costs

A main type of cost depends on the empty freight car mileage. The cost includes, for example, energy costs and freight car wear costs. It is strongly related to the transportation distances and can be modeled as proportional to the number of moved cars and the transportation distances. This distance related cost can be influenced in the tactical planning stage by the design of the timetable, and in the operational planning stage by changing the routes of the freight cars. In the distribution of empty cars this type of cost is of special interest, since the origin-destination flows of empty cars are not given, but can be influenced in the operational planning.

In the optimization model we include a cost that is proportional to the number of cars transported in each origin-destination pair. This cost is also directly related to transportation distance and we refer to this cost as the transportation cost.

3.3 Yard costs

For costs associated to classification yards, there is an important distinction between costs associated to strategic and tactical decisions on one hand, and on operational costs on the other hand. The strategic and tactical phases of the planning process determine the existence of the classification yards, which is certainly associated to a large cost. In the tactical planning it is determined which blocks are built at each yard, and in the operational planning it is determined how the blocks are built. Consequently, at the operational planning level the only costs of interest are the costs that depend on the work that has to be done in the yard. The operational costs of the yard are connected to the freight car handling that is performed in the yard. In the optimization model we do not include any strategic or tactical yard costs, but we include a cost representing the operational yard costs for freight car handling. This is discussed below.

The objective of the yard operations is to build blocks of cars to form the outgoing trains. The yard operations are very complex, and it is difficult to make a correct and detailed analysis of the costs that are associated with the yard operations. In order to analyze the cost at the yards, we define the term *car-cluster*, which is related to the more

common term "block".

- A *block* is a group of cars sent between two yards. A block is moving in trains and through yards without any manipulation on the cars in the block. The origin yard of the block does not have to be the origin of all cars in the block, and the destination of the block does not have to be the final destination of all cars in the block. Thus, cars travel in a block in a *part* of their transport. Which blocks that can (or have to) be built on a yard depend on the design of the yard at the strategic planning level and on the design of the timetable at the tactical level.

- A *car-cluster* is a group of cars that has a common initial yard and a common final destination yard and that are moving together in trains and through yards during the *whole* transport. Cars in a car-cluster are physically connected all the way from the origin to the destination, and they never have to be separated during the transport. Thus, one car-cluster can be considered as one handling unit in the yard operations, causing one unit of cost. The number of car-clusters is determined at the operational planning level and is given by the transportation pattern. Since trains are defined to be time-and-day-specific (each train exists only once), each car-cluster exists only once. The cars in the car-cluster can be loaded or unloaded.

When a train arrives at a yard, the freight cars are classified with respect to outgoing trains. The incoming trains often have to be split into many small car groups in order to build the outgoing trains. In the worst case, each car group consists of one car-cluster.

The operational costs at a yard can then be described as consisting of two components:

- One car-handling cost that is proportional to the number of cars handled at the yard.

- One group-handling cost that is proportional to the number of car groups that are handled at the yard.

In the model presented in this paper, yard costs are modeled as follows. The car-handling cost is determined by the route of the car. The sum of the car-handling costs related to all cars on a given route is included in the transportation cost (that is thus proportional to the number of cars moved on that route). For the group-handling cost, we approximate the number of car groups handled at a yard by the number of car-clusters that pass through the yard. It is an approximation because when two car-clusters use the same sequence of trains trains for part of their routes, the car-clusters can be consolidated for their common itinerary and avoid being disconnected at intermediate yards. In this case, the number of car handling operations will not be strictly equal to the number of

car-clusters. Differences are not such, however, as to prevent the number of car-clusters from being a good base for the evaluation of the yard effort and cost [10]. Consequently, a (fixed) cost is associated to each car-cluster, equal to the sum of the group-handling costs of all the yards the car-cluster passes through.

In railway systems that are not strictly scheduled, handling times at classification yards are of major concern. In these cases, it should be even more interesting to find transportation patterns that reduce the yard workloads. Handling times often depend nonlinearly on workloads, however. Therefore, for such railway systems, another price model should be used. For the strictly scheduled railways we consider in this paper, the cost model we propose is nevertheless appropriate.

4 Mathematical model

In this section we specify the modeling assumptions and present a mathematical model for the problem studied in this paper. The objective of the model is to propose origin-destination specific movements of empty freight cars, and also to determine their routes in the railway system.

4.1 Model specification and assumptions

In the strictly scheduled railway system, the timetable is determined at the tactical planning stage. Consequently, arrival and departure times of trains are fixed, available capacities of the trains are known, and the possible routes in the railway network are stipulated. In a strictly scheduled railway, the transportation times and handling times at the classification yards are given by the design of the timetable. (Timetables are designed to avoid congestion-generated delays at classification yards.) The timetable does not have to be cyclic, which means that the possible movements may differ from period to period. The reservation possibilities for freight cars in the trains certify that the planned transports can be performed.

There exist many types of freight cars. We assume that they all have the same size (length and weight), which means that the transportation capacities of the trains can be expressed by the number of empty freight cars. This assumption is easily generalized by modifying the capacity constraints in the model, which is further discussed below.

The loaded transports are a given input to the empty distribution planning, either known or given by a forecast. As a consequence, the supply and demand for each car type at each terminal are known. Each demand can only be fulfilled by one car type. We

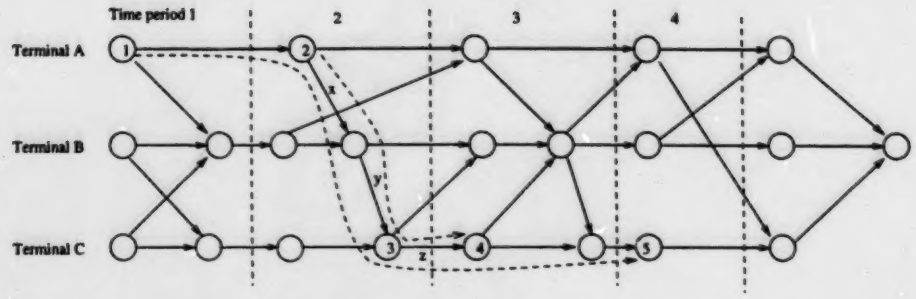

Figure 2: The train network and two paths

assume that there always exists a way to fulfill all demands and that a demand cannot be backordered (the demand that are possible to fulfill can be calculated in a preprocess before applying the model, for example in the way described in [8]). The demands for loaded transports are known before the empty distribution is planned, which means that we can use a deterministic approach to the problem. Further, the loaded transports are considered as unalterable, and the degree of freedom in the planning only relates to how to distribute the flow of the empty freight cars. The supply, demand, transportation possibilities and transportation capacities may vary between the days in the planning period, so the distribution problem is represented in a time-expanded model.

The model formulation is based on a time-expanded network, that is referred to as the *train network*, illustrated in Figure 2. The flow in the network represents *empty* freight car movements, and the flow consists of several commodities, where each commodity represents one freight car type. Each row of nodes corresponds to one railway terminal. The planning period is divided into time periods (days). Each terminal is represented by at least one node in each time period. Additional nodes may be needed in a time period to represent the possible connections between arriving and departing trains. We assume that the supply and demand of empty cars are realized at the beginning of each time period. The balance of the node at the beginning of each time period represents the net supply or demand for cars of each respective type. Note that each commodity may have several sources and sinks and that one node can be a source and a sink of several commodities. An arc corresponds either to an inventory within or between two time periods, or to a train between two terminals. An arc representing a train has a "bundle" capacity constraint on the total number of empty freight cars that can be transported in the train. We assume that the trains' capacities are given as the maximum number of empty cars that can be transported in each train.

As defined in Section 3.3, a car-cluster is a set of cars that can be held together all the way from the initial origin to the final destination, which means that the cars in a car-cluster are transported on the same day in the same trains. We assume that all freight cars, loaded or empty, that have the same initial origin and final destination and are to

be transported in the same set of trains all the way from the origin to the destination, are grouped into one car-cluster. This means that each car-cluster can be characterized by a unique set of trains that are used for the transport, and we define the set H_q to be the set of trains that are used for transporting car-cluster q.

A path in the train network consists of a sequence of arcs from a source node to a sink node. In Figure 2, two paths are depicted with dashed lines. One path originates from node 1 representing terminal A in time period 1 and terminates at node 5, representing terminal C in time period 4. The other path originates from node 2, representing terminal A in time period 2, and terminates at node 4, representing terminal C in time period 3. Both these paths include the arcs x and y, representing trains from A to B and from B to C, respectively. The flows on these two paths both represent transports of empty cars from A to C, and in fact the same trains are used to perform the transports. That is to say, both paths represent the same spatial movement, the origin and destination terminals are the same and the same set of trains that are used to transport the empty cars. The difference in interpretation between the two paths is the temporal aspect of when the cars become available (different source nodes) and when cars are used to fulfill demand (different sink nodes). Since the cars that are represented by flows on the two paths have the same origin and, destination terminals are transported in the same trains, these cars constitute one car-cluster. Let this car-cluster be denoted q' and let $H_{q'}$ be the trains used to transport q'.

Consider all the paths in the train network in Figure 2 representing movements from one terminal to another. These paths can be divided into a number of subsets, each one including the same set of train arcs. In Figure 3, all paths in the subset that represent transport in the trains in $H_{q'}$ are depicted. All these paths represent transportation of the car-cluster q', and we denote this subset of paths $P_{q'}$.

All paths in $P_{q'}$ have three arcs in common, in Figure 3 depicted with bold links. The arcs that are in common correspond to the trains in $H_{q'}$ and one additional inventory arc. In Figure 2 these arcs are the train arcs denoted x and y, and the inventory arc denoted z. (Demand is realized in the beginning of the time periods, therefore node 3 cannot be a sink node and no path can terminate at node 3.)

For each subset of paths P_q that corresponds to a unique set of trains H_q and consequently includes a unique set of train arcs, we define a *kernel path* p_q as consisting of all the arcs (train arcs and inventory arcs) that are in common for all paths in the set. In Figure 3, the kernel path $p_{q'}$, corresponding to the subset $P_{q'}$, is represented by the chain of bold arcs. A kernel path p_q differs from the paths in its corresponding set P_q of "original" paths by the fact that it does not include inventory arcs before the beginning of the time period of the departure of the first train of H_q, nor of any inventory arcs after the beginning of the time period of the arrival of the last train of H_q. Each kernel path p_q has a unique origin terminal, destination terminal, departure time, arrival time and

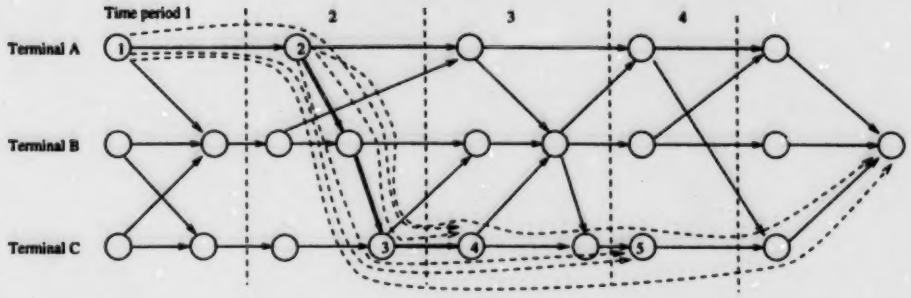

Figure 3: The train network and kernel paths

corresponds to a unique set of trains, H_q. In Figure 3, all paths in the subset of paths in $P_{q'}$ that correspond to the kernel path $p_{q'}$, are illustrated with dashed lines.

A kernel path represents one way of performing a transport between two terminals and therefore the kernel path represents exactly one car-cluster. All kernel paths emanate and terminate in the beginning of the time periods, since the supply and demand are assumed to be realized in the beginning of the time periods. The flow on a kernel path p_q corresponds to the total flow, of any commodity, on the paths in P_q. Consequently, the flow on a kernel path can originate from different supply nodes representing the same supply terminal in different time periods, and can terminate at different sink nodes, representing the same demand terminal in different time periods. A kernel path has no capacity restriction *per se*, instead there are capacities on the train arcs that are included in the kernel path. The capacity constraint is a bundle constraint over the sum of the flows of all kernel paths that include a given train arc.

In the model we include a cost for sending a car-cluster, which is equivalent to assigning a fixed cost to each kernel path which has a nonzero flow. This means that the flows on several paths in one subset P_q generate only one fixed cost, since the flows correspond to exactly one car-cluster q. For example, the flow on all the depicted paths in Figure 3 generate one fixed car-cluster cost, since the paths correspond to exactly one kernel path and the freight cars corresponding to the flow on the paths are assumed to be connected into one car-cluster.

In Figure 4, all possible kernel paths of the train network in Figure 2 are indicated. The possible kernel paths are given by the transportation structure of the railway system. In the railway systems we study in this paper, we consider the kernel paths as generated beforehand, defined by the timetable. It is possible to include a column generation approach to generate kernel paths during the solution of the optimization model, but this is not further studied in this paper.

The flow assigned to a kernel path can be directly interpreted in terms of the number

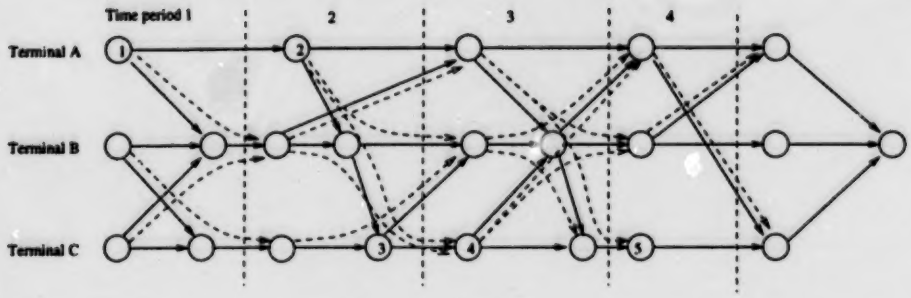

Figure 4: Train Network with all kernel paths

of empty freight cars of each type to be transported in one car-cluster between two terminals in a specific combination of trains.

We assume a finite length of the planning period. The model should be used using a rolling horizon concept. The solution of the optimization model defines a distribution pattern for the whole planning period, but only the part corresponding to the first time period is implemented. As time goes by and new information becomes available the planning period is shifted forward to include a new time period. To compensate for effects of the finite planning period we can use a salvage value for empty cars in the final time period. The salvage value at the different terminals depends on how useful an empty freight car is at the terminal. The end-of-horizon effects can be handled in different ways, see [9] for a thorough analysis.

Given the definition and enumeration of kernel paths, the train network, illustrated in Figure 4, is transformed into another time expanded network, denoted the *kernel network*,illustrated in Figure 5. Each kernel path in the train network in Figure 4 is transformed into a kernel arc between the origin and destination terminals in the kernel network of Figure 5. In the kernel network there is only one node for each terminal and time period and the possible trains connections are implicitly defined by the kernel arcs. Therefore, in the kernel network there is no need to define more than one node fore each terminal in each time period. On the kernel arcs there are variable flow costs representing the per car transportation cost. In addition, there is a fixed cost on the kernel arcs which appears if a nonzero flow is assigned to the kernel arc. If desirable, it is easy to include a commodity specific inventory cost on the inventory arcs in the kernel network. The capacity restriction of a kernel arc in the kernel network correspond to the capacity restriction of a kernel path in the train network. Thus, for each train there is a side constraint which limit the total flow on all kernel arcs that include this train. Hence, each capacity constraint limit the flow on several kernel arcs, since several kernel arc may correspond to movement in the same train. Moreover, each kernel arc can be included in several capacity constraints, since each kernel arc can correspond to movements in several trains.

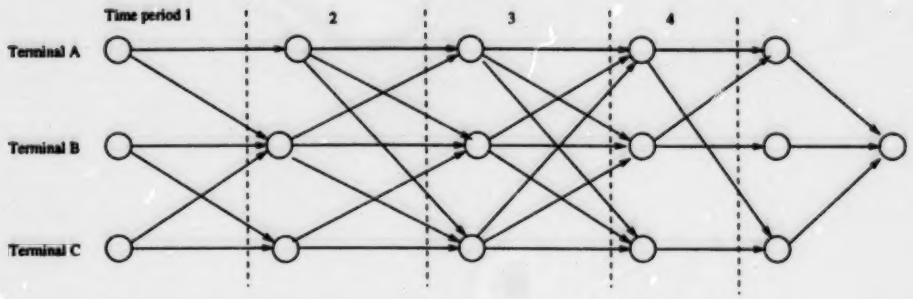

Figure 5: Kernel network corresponding to kernel paths in Figure 4

In the kernel network, the kernel arcs can be regarded as design arcs, and the optimization includes determining which of these arcs that shall be included in the solution. Hence, the model can be interpreted as a special case of a multicommodity capacitated network design model. The difference is that in a traditional network design model (see, e.g., [11]), capacity constraints are defined for the total flow on each respective design arc. In our model, each capacity constraint represents a limit on the number of freight cars transported in one train. Therefore, each capacity constraint limits the total flow on *several* design arcs, since flow on many different kernel arcs can represent transports in the same train. Furthermore, in the kernel network each design arc can be included in several capacity constraints as a result of the fact that the kernel arcs can represent transports in several trains.

4.2 Model formulation

In this section we define all variables and input data and present the mathematical formulation of the model. The formulation is based on the kernel network illustrated in Figure 5. In the model formulation we refer to kernel arcs in the kernel network, but these can alternatively be interpreted as kernel paths in the train network.

General definitions

N: The set of terminals in the railway network.

T: The number of time periods within the planning period.

K: The set of freight car types.

G: The set of trains.

b_{nt}^k: Net supply (i.e., supply minus demand) of cars of type k at terminal n in period t. The net supply is negative if the demand is greater than the supply, $k \in K$, $t = 1, 2, \ldots, T+1$, $(b_{n,T+1}^k = 0)$.

i_{n0}^k: Inventory of empty freight cars of type k at terminal n in the beginning of the planning period, $n \in N$, $k \in K$.

Path definitions

P: The set of all kernel arcs in the network.

P_{nt}^o: The set of all kernel arcs emanating from terminal n in period t, $P_{nt}^o \subset P$, $n \in N$, $t = 1, 2, \ldots, T+1$, $(P_{n,T+1}^o = \emptyset)$.

P_{nt}^d: The set of all kernel arcs terminating at terminal n in period t, $P_{nt}^d \subset P$, $\forall n \in N$, $t = 1, \ldots, T+1$, $(P_{n1}^d = \emptyset)$.

P_g: The set of all kernel arcs that include transport in train g, $g \in G$, $P_g \subset P$.

m_p: Maximum number of empty cars that can be sent in the car-cluster corresponding to kernel path p, $p \in P$,.

u_g: Transportation capacity (in number of freight cars) of train g, $g \in G$.

Costs

f_p: The cost for sending a car-cluster from an origin terminal to a destination terminal in the trains corresponding to kernel arc p, $p \in P$.

c_p: The variable cost for sending a car from an origin terminal to a destination terminal in the trains corresponding to kernel arc p, $p \in P$.

v_n^k: Salvage value of a car of type k at terminal n at the end of the time horizon, $n \in N$, $k \in K$.

Decision variables

x_p^k: The number of empty cars of type k transported in the trains corresponding to kernel arc p, $p \in P$, $k \in K$.

y_p: $\begin{cases} 1 \text{ if any cars are sent in the car-cluster corresponding to kernel arc } p,\ p \in P; \\ 0 \text{ otherwise.} \end{cases}$

i_{nt}^k: The number of cars of type k in inventory at terminal n at the end of period t, $n \in N$, $k \in K$, $t = 1, 2, \ldots, T+1$.

Formulation

$$F^* = \min \sum_{p \in P} \left(f_p y_p + \sum_{k \in K} c_p x_p^k \right) - \sum_{n \in N} \sum_{k \in K} v_n^k i_{n,T+1}^k \tag{1}$$

Subject to

$$i_{nt}^k = i_{n,t-1}^k + b_{nt}^k + \sum_{p \in P_{nt}^d} x_p^k - \sum_{p \in P_{nt}^o} x_p^k \; \forall k \in K, \; \forall n \in N, \; t = 1, 2, \ldots, T+1 \tag{2}$$

$$\sum_{k \in K} x_p^k \leq m_p y_p, \quad \forall p \in P, \tag{3}$$

$$\sum_{p \in P_g} \sum_{k \in K} x_p^k \leq u_g, \quad \forall g \in G, \tag{4}$$

$$x_p^k \geq 0, \quad \forall k \in K, \; \forall p \in P, \tag{5}$$

$$y_p \in \{0, 1\}, \quad \forall p \in P, \tag{6}$$

$$i_{nt}^k \geq 0 < \quad \forall k \in K, \; \forall n \in N, \; t = 1, 2, \ldots, T+1. \tag{7}$$

Description of the Objective Function and the Constraints

The objective is to minimize the total costs of the distribution process. The objective function (1) has two parts. The first part represents the car-cluster costs for utilizing kernel arcs and the variable transportation costs for assigning flow to the kernel arcs (i.e., costs depending on the number of cars sent in a car-cluster). The second part of the objective function is the costs associated with the time horizon and represents of the salvage values for cars at different terminals.

The constraints (2) are the inventory balances, stating that the number of cars in inventory of each car type at each terminal is equal to the outgoing inventory of the previous time period plus supply of cars plus cars arriving from other terminals minus cars sent to other terminals. The constraints (2) correspond to the node balances of the kernel network interpretation of the model.

Constraints (3) state that the binary variable y_p has to be one if any flow is assigned to the kernel arc p (i.e., that a car-cluster exists if any freight cars are sent in the corresponding OD-pair). Each constraint of type (4) bound the total flow of all commodities on all kernel arcs that correspond to flow on one constrained arc in the train network, (i.e., the number of cars in all car-clusters in a train must not exceed the capacity of the train). If the model should represent that freight cars of different types have different length and weight, constraints (4) could be reformulated and duplicated; one set constraints for length capacities and set of constraints for weight capacities. The difference

between this model and a generic network design model (as presented in [11]) is constraints (4). Considering the y_p variables as design variables connected to design arcs and the x_k^p variables as the flow on the design arcs, constraints (4) states that there is a limit on the total flow on sets of design arcs.

In principle, all variables should be integer valued, since the solution is interpreted as an empty flow pattern. However, in order to simplify the solution procedure, we relax the integrality constraints on flow variables but include integrality constraints on the binary variables connected to the design arcs (y-variables). Nevertheless, numerical experiments with the solution procedure we propose in the next section has shown to assign integer values to all variables in almost all test cases. If the model is used to study principles of distribution performance, as in this paper, or for strategic or tactical planning, a non-integer valued solution is probably as useful as an integer valued solution. On the other hand, if the model is used for operational planning, the integrality constraints of the flow variables should not be relaxed and the applied solution method should take the integrality of the flow variables into account.

5 Solution method

In this section, we present a tabu search heuristic [6] for solving the proposed empty freight car distribution model described in Section 4. Network design problems are known to be difficult to solve ([11]), which motivates the use of a metaheuristic solution approach to our special case of the network design problem. More details about the tabu heuristic can be found in [3]. The tabu search procedure is summarized in Table 1.

A solution in the tabu-search procedure is defined by a binary vector that determines the values of the y-variables, i.e., the variables that determines whether each kernel arc is open or closed. A solution is evaluated by solving the subproblem $P(\bar{y})$ that results from fixing all binary variables to $\bar{y}$. $P(\bar{y})$ is a linear multicommodity network flow problem that is solved by an LP-solver. The objective function value of $P(\bar{y})$ is denoted $z(\bar{y})$ and includes all flow costs (transportation costs) but does not include the design costs of the kernel arcs (car-cluster costs). The value $F(\bar{y})$ of solution $\bar{y}$ in the tabu search is obtained as $z(\bar{y})$ plus the costs of the binary variables $\bar{y}$ that are fixed to one.

We define two types of neighbourhoods to a solution $\bar{y}$. The neighbourhood of type $N_1(\bar{y})$ has one additional kernel arc closed, and the neighbourhood of type $N_2(\bar{y})$ has one additional kernel arc open. Hence, a move in the search procedure is either to open one closed kernel arc p or to close one open kernel arc p, i.e. to change the value of exactly one binary variable y_p. Neighbours are evaluated by resolving the multicommodity flow subproblems with the appropriate changes in the fixed binary variables.

We only move to a neighbour y' of type $N_2(\bar{y})$ if $z(y')$ is strictly better than $z(\bar{y})$. The reason for this is that a neighbour of type $N_2(\bar{y})$ can never have a worse $z(y')$ than $z(\bar{y})$. If $z(y')$ is not improved when moving to the neighbour y' in the neighbourhood $N_2(\bar{y})$, it means that the kernel arc that is opened when moving to y' is not used to improve $z(y')$, and therefore there is no incentive to move to y'.

Generally, the number of kernel arcs is very large while only a small portion of the kernel arcs have a nonzero flow. Hence, the neighbourhood $N_1(\bar{y})$ is small while the neighbourhood $N_2(\bar{y})$ is very large. Therefore it is intractable to calculate the exact value of each neighbour in $N_2(\bar{y})$. Instead, we first calculate a target value (estimated value) of the change in the objective function value when moving to a neighbour in $N_2(\bar{y})$. The neighbours with the best target values are then evaluated exactly.

The target value for opening a kernel arc can be calculated in many ways. We tried three ways of calculating the target value, all of them are based on the dual variables obtained from the solution of the linear subproblem $P(\bar{y})$. The target value $\bar{f}_p$ for the kernel arc p, emanating from node i and terminating at node j, is calculated in the following ways:

1. Reduced cost of the y_p-variables in the LP-relaxation of model: $\bar{f}_p = f_p + m_p(\min\{0, \bar{c}_p - \sum_{g \in P_g} \delta_g)$ where
   $\bar{c}_p = c_p - \max_k\{d_j^k - d_i^k\}$, i.e., the arc flow cost minus the maximum difference in node prices between node j and node i for the different commodities;
   d_i^k: node price of node i for commodity k, i.e., dual variable of the appropriate constraint (2);
   δ: dual variable of the constraints (4);
   c_p, f_p, m_p: same as definitions as above.
   m_p can be considered as an estimate of the flow on the kernel arc if it is opened. Since m_p is equal to the capacity of the arc, it tends to overestimate the flow on the arc when it is opened.

2. Modifying target value of type 1 by replacing m_p with an estimate the flow of each commodity if the kernel arc is opened. The flow is estimated as minimum of inflow of node i and outflow of node j.

3. Using $\bar{f}_p = f_p + \bar{c}_p$ as target value. This correspond to an estimate of the change in the objective function value if one unit of flow is assigned to the opened kernel path.

In the calibration tests, target values of type 3 gave the best results and is used in this paper. Possibly, the estimates of the flow on the opened kernel arc used in the other target value formulas are not better than always assuming that the flow will be one unit.

In order to speed up tabu search, the neighbourhood $N_1(\bar{y})$ is not completely evaluated, but a random sample of the neighbours in $N_1(\bar{y})$ is evaluated. The move is made to the best evaluated neighbour of type $N_1(\bar{y})$ or $N_2(\bar{y})$.

There are two different short time tabu lists. One list contains the kernel arcs that were recently opened; the tabu restriction is not to close these kernel arcs. The other list contains the kernel arcs that were recently closed, which are tabu to open. The tabu restriction is overridden if it leads to a new best solution.

Artificial arcs are added to the model so that the problem always has a solution given any set of open kernel arcs. The artificial arcs are added from the super sink beyond the time horizon to the first node in the beginning of the planning period of each terminal. In this way we can utilize "tunneling" in the heuristic, i.e., we can move from one feasible solution to another feasible solution (feasible with respect to the original problem) via a set of infeasible solutions. During the tabu search, the cost of the artificial arcs are regularly lowered and raised, in order to utilize tunneling and to provide feasible solutions, respectively.

We tried two strategies for updating the cost of the artificial arc costs. In both strategies, artificial arcs costs are updated every i-th iteration.

1. All artificial arc costs are multiplied with 2 if the solution has been infeasible in all the i last iterations, and the all artificial arc costs are divided by 2 if the solution has been feasible in all the i last iterations. This strategy was proposed by [5].
2. Modifying strategy 1 by multiplying the cost on one individual artificial arc with a growth factor α ($\alpha > 1$) if there has been any flow on that arc in any of the i last iterations. The cost on all artificial arcs are multiplied with a decrease factor β ($\beta < 1$) if the solution in all the last i iterations has been feasible.

In the calibration tests, strategy 1 resulted in frequent but small improvements in the currently best solution, while strategy 2 resulted in less frequent updates of the currently best solutions, but then with larger improvement steps. This is a result of the finer tuning of the artificial arc costs in strategy 2, which allows many ways of tunneling, making it slower to find a feasible solution, but the feasible solutions achieved after a tunneling sequence are often of good quality. After one hour of computation time in the tests, the solutions provided by strategy 2 were often better than by strategy 1, and strategy 2 is used in this paper.

An initial solution to the problem is found by first solving the problem with all fixed costs set to zero, i.e., the variable transportation costs are minimized. All kernel arcs which have a nonzero flow are then considered as open. This solution is improved by closing the open kernel arcs one by one. If the optimal variable cost is not changed by

the closure of one kernel arc, this kernel arc is kept closed; otherwise the kernel arc is reopened.

More details about the numerical experiments regarding the specification of the tabu heuristic can be found in [3].

The tabu search algorithm is initiated with the following set of parameters:
v_1, v_2: Number of neighbours to examine in $N_1(\bar{y})$ and $N_2(\bar{y})$, respectively.
l_{open}, l_{close}: Number of iterations it is declared tabu to close an opened arc and to open a closed arc, respectively.
i: Number of iterations between artificial arc cost updates.
α, β: Growth and decrease factors for artificial arc costs, respectively.

Step 0: Select a feasible starting solution $\bar{y}$. Set the currently best solution $y^* := \bar{y}$. Initialize iteration counter $t := 1$.
Step 1: Randomly select v_1 neighbours in $N_1(\bar{y})$. Calculate the corresponding objective function values F of the neighbours.
Step 2: Calculate the target value of all neighbours in $N_2(\bar{y})$. Calculate the objective function value F of the v_2 neighbours with the best target values.
Step 3: Sort all the evaluated neighbours y into a list L according to $F(y)$, best neighbour first.
Step 4: Pick the first neighbour y' in L.
Step 5: If y' is tabu: if y' correspond to an infeasible solution, or $F(y') \geq F(y^*)$, remove y' from L and go to step 4.
Step 6: If y' is a neighbour of type $N_2(\bar{y})$ and $z(y') \geq z(\bar{y})$, remove y' from L and go to step 4.
Step 7: If y' is a feasible solution and $F(y') < F(y^*)$, set $y^* := y'$.
Step 8: Update tabu lists according to the opened or closed kernel arc in y', and by (potentially) making some kernel arc non tabu.
Step 9: If t is a multiple of i, update artificial arc costs.
Step 10: Move to the neighbour y', set $\bar{y} := y'$, increase iteration counter $t := t + 1$, go to step 1.

Table 1: The tabu search procedure

6 Computational results

The objective of the numerical experiments is to show that problems of real-life size are solvable in reasonable time. Also, we show that the distribution plans obtained with the model utilize economies of scale.

Numerical experiments were performed on a set of randomly generated test problems. The test problems are constructed so that the structure is closely similar to the situation at Swedish Railways. We assume that trains run in a hub-and-spoke structure, corresponding to major classification yards and terminals. All terminals are assigned to exactly one major classification yard and there are one to three daily trains between a classification yard and its terminals. Between the major classification yards there are also one to three daily trains. Capacities of trains are chosen so that transportation capacities are reasonably scarce. The capacities of the trains are randomly generated around a day specific mean value. For each freight car type, each terminal is first assigned to be a demand or supply terminal of different levels, and then supply and demand are randomly generated according to these base levels. The salvage value for empty freight cars at the end of the planning period is set to zero in all tests presented in this paper. The sizes of the test problems can be found in Table 2. Test problem $p6$ approximately corresponds to the size of the Swedish railway network. Problem $p7$ is a network without the hub-and-spoke structure ("only hubs and no spokes").

Prob	Hubs	Terminals	Commodities	Time periods	Kernel arcs 0/1 variables	Continuous variables	Constraints
p1	3	6	2	3	262	654	643
p2	4	12	5	4	728	4360	4117
p3	6	6	10	5	1288	14020	13603
p4	6	25	10	5	12061	124670	122867
p5	15	25	15	5	21701	349841	329305
p6	5	50	20	5	18084	386764	369250
p7	25	25	15	4	2079694	33277354	31200075

Table 2: Sizes of test problems

The tabu search algorithm is implemented in the programming language C and CPLEX (version 4) [2] is used as LP-solver in the search algorithm. Tests are performed on a multiuser SUN Ultrasparc 2/2200 with 2 GB RAM. To evaluate the performance of the tabu-search algorithm, we also solved the test problems with the mixed integer programming (MIP) solver CPLEX (version 6.5).

Experiments with different parameter settings are presented in [3]. Table 3 summarize the evaluated and the best values (column "Best") for the parameters of the procedure. To calibrate the parameters, three test problems of about the same size as problem $p5$ were used.

The first set of tests illustrates the performance of the tabu heuristic in comparison to CPLEX. The results are presented in Table 4. The tabu search algorithm was stopped after one hour. CPLEX was stopped after 10 hours CPU time or when 300000 nodes have been searched or CPLEX ran out of memory. For the smaller problems ($p1$, $p2$, $p3$, and $p4$), CPLEX found integer valued solutions, but could not prove optimality. For

Parameter	Tested values	Best
Length of tabu list for opened kernel arcs	5, 10, 15	10
Length of tabu list for closed kernel arcs	5, 10, 20	10
Number of evaluated neighbours in $N_1(\bar{y})$	25, 50, 100, all	50
Number of evaluated neighbours in $N_2(\bar{y})$	25, 50, 100, 200	50
Move to a neighbour y' of type $N_2(\bar{y})$ if $z(y') \geq z(\bar{y})$	yes, no	no
Number of iterations i between artificial arc cost updates	1, 2, 4, 8	4
Growth factor for artificial arc costs	1.1, 1.2, 1.3	1.2
Decrease factor for artificial arc costs	0.9, 0.8, 0.7	0.7

Table 3: Tested and best parameter settings.

the smallest problem ($p1$), the solution found by the tabu heuristic was as good as the solution found by CPLEX, but for all other problems, the solution found by the tabu heuristic was better. For problems $p5$ and $p6$ CPLEX found no integer valued solutions in 10 hours of computing time. For problem $p7$, not enough memory could be allocated to load it into CPLEX.

In all test problems, the gap between the best found integer valued solution and the LP-relaxation is large. For the smaller problems ($p1$, $p2$, $p3$), the LP-relaxation is solved quickly by CPLEX. However, there is no straightforward way of achieving an good integer valued solution from a solution from an LP-relaxation since it is difficult to round the fractional variables so that both the node constraints and capacity constraints remain satisfied.

The tabu search heuristic is much faster than CPLEX, which can be verified by comparing the time it takes for the tabu search to find a solution comparable with the best solution from CPLEX. For problem $p1$, the tabu search found the best solution in 2 seconds while CPLEX used 25 seconds to find an equally good solution. Moreover, CPLEX needed more than 9 hours to find its the best solution in problem $p2$, while the tabu search found a better solution in less than 1 minute. For problems $p2$, $p3$, $p4$, and $p6$ the best solution that the tabu-search heuristic could find was achieved after almost one hour, that is, just before the calculations were stopped, which indicates that the algorithm was still progressing.

Our conclusion is that the tabu search algorithm can be used to solve the proposed optimization model for problem instances of realistic size. The optimizing algorithm in CPLEX cannot prove optimality even for the smallest problems, and with larger problems the best solution found by CPLEX is inferior to solutions found by the tabu search. In addition, tabu search is faster.

As mentioned above, the x-variables should in principle be integer-valued, but this constraint is relaxed. However, for the test problems, the tabu search algorithm yielded

Problem	Tabu solution	CPLEX MIP UBD	LP-relaxation
p1	19208	19208	17004
p2	31501	32457	24442
p3	136809	137474	126321
p4	683778	696249	620556
p5	305933	-	276454
p6	414617	-	328063
p7	149490	-	-

Table 4: Performance comparisons

integer-valued solutions for all variables, in most steps of the search. On average, for the test problems, 88% of the visited solutions are integer-valued, and 93% of the solutions that were recorded as "currently best solution" are integers. Of the non-integer valued solutions, 80% correspond to infeasible solutions (i.e., there is flow on some artificial arc). In the non-integer valued solutions, on average 3% of the variables currently used in the tabu search are not integer-valued (variables corresponding to closed kernel arcs are not included in the set of currently used of variables).

In order to guarantee that the tabu search procedure results in an integer-valued solution, the method can be modified to only store solutions with all integer valued variables as "currently best solution". Since most visited solutions are integer-valued, there is a good chance that this simple modification will result in good integer solutions.

The second test illustrates the impact on the empty freight car distribution of the introduction of the costs for transporting a car-cluster. Table 5 shows how the number of car-clusters in the best found solutions depends on the size of the car-cluster cost. (The optimal transportation cost when no costs on car-clusters are considered is 280232.) The results correspond to the mean values of three test problems after one hour of calculation time, during which about 250 iterations of the tabu heuristic were performed. All the three test problems have about the same sizes as problem p5 in Table 2.

The objective of introducing a cost on car-clusters in the optimization model is to reduce the number of car-clusters that are transported in the railway system. The results in Table 5 indicate that the introducing car-cluster costs into the formulation has a significant effect on the solutions. As the cost for each car-cluster is increased, the number of car-clusters transported in the empty freight car distribution is reduced. On the other hand, the transportation cost increases. Table 5 shows the number of car-clusters used in the best found solution, as well as the increase of the transportation cost compared to a solution in which costs on car-clusters are not included. The results show that if the cost for sending a car-cluster is increased, the transportation pattern is organized so that fewer car-clusters are used while the transportation cost increases slightly. Pre-processing

the problems before the tabu search algorithm was applied, reduced the number of car-clusters used in the solutions from 309 to 165, while the transportation cost was not increased at all.

Cost per car-cluster	Number of car-clusters	Increase in transportation cost
0	309	0
50	130	774 (0.3%)
100	123	1254 (0.4%)
200	112	3459 (1.2%)
400	101	5874 (2.1%)
800	98	8694 (3.1%)

Table 5: Impact of car-cluster transportation costs

We also made experiments with non-zero salvage values. No major differences in results and conclusions were found. The performance of the tabu heuristic is not influenced by the salvage values. The number of transported car-clusters and the transportation cost are slightly increased as a consequence of the preferences about the transportation pattern that are stated by the salvage values, but the general conclusions stated above are still valid.

7 Conclusion

In this paper we have formulated a model for operational empty freight car distribution that includes explicit consideration of economies of scale. In order to capture economies of scale we include a cost for transporting a car-cluster between an origin and a destination in the railway system. The optimization model is based on a time-expanded network where the cost on a car-cluster is represented as a fixed cost for assigning any flow to a path. The optimization model is a special case of a capacitated network design model with the unusual feature that the capacity restrictions limit the total flow on several arcs. A tabu search metaheuristic is used to solve the model. The heuristic successfully finds solutions to the model that are of a much better quality than those a MIP-solver could find.

The solutions obtained from the model display the economies of scale effect, in the sense that the car movements in the solution are coordinated to reduce the number of car-clusters that are transported in the railway network. For instance, in our tests, the number of car-clusters transported to fulfill the supply and demand in the empty freight car distribution could be reduced by 67%, while the transportation cost increased by only 2%

A key characteristic of the model is the definition of kernel paths and the introduction of fixed costs on such paths. This new modeling approach can certainly have applications in other problem areas as well. The important aspect the model captures is the economies of scale associated to transporting more than one unit from an origin to a destination, while capacity restrictions limit the amount that can be transported. The capacity restrictions are not applicable directly to the origin-destination transports, but to sublinks that together form the transports.

The model of empty freight car distribution can be further enhanced by, for example, considering coordination between different car-clusters. Even though the tabu heuristic used in this paper is rather simple, we achieved promising results. A more refined tabu search strategy should permit to further improve these results.

Acknowledgments

This work was financed by Swedish Railways (SJ). Partial funding for this project has also been provided by the Natural Sciences and Engineering Council of Canada and by the Fonds F.C.A.R. of the Province of Québec.

References

[1] J.-F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science* 32(4), 380–404, 1998.

[2] CPLEX Optimization, Inc.. *Using the CPLEX Callable Library*, 1994.

[3] T. G. Crainic, M. Gendreau, K. Holmberg, M. Joborn, J. T. Lundgren, and K. Holmberg. A tabu heuristic for the network design problem with fixed costs on paths. Working paper LiTH-MAT-WP-1997-01, Department of Mathematics, Linköping University, Sweden 1997.

[4] P. J. Dejax and T. G. Crainic, A review of empty flows and fleet management models in freight transportation. *Transportation Science* 21(4), 227–247, 1987.

[5] M. Gendreau, A. Hertz, and G. Laporte. A tabu search heuristic for the vehicle routing problem. *Management Science* 40, 1276–1290, 1994.

[6] F. Glover and M. Laguna. *Tabu search*, Kluwer, Norwell, MA, 1997.

[7] A. E. Haghani. Formulation and solution of a combined train routing and makeup, and empty car distribution model. *Transportation Research B* 23, 433–452, 1989.

[8] K. Holmberg, M. Joborn, and J. T. Lundgren. Improved empty freight car distribution. *Transportation Science* 32, 163–173, 1998.

[9] R. E. Hughes and W. B. Powell. Mitigating end effects in the dynamic vehicle allocation model. *Management Science* 34, 859–879, 1988.

[10] M. Joborn. Empty freight car distribution atSwedish Railways - analysis and optimization modeling. Licenciate Thesis, LiU-TEK-LIC-1995:31, Linköping University, Sweden, 1995.

[11] T. L. Magnanti and R. T. Wong. Network design and transportation planning: Models and algorithms. *Transportation Science* 18(1), 1–55, 1984.

[12] W. B. Powell and T. A. Carvalho. Dynamic control of logistics queueing networks for large-scale fleet management. *Transportation Science* 32(2), 90–109, 1998a.

[13] W. B. Powell and T. A. Carvalho. Real-time optimization of containers and flatcars for intermodal operations. *Transportation Science* 32(2), 110–126, 1998b.

[14] M. A. Turnquist. Economies of scale and network optimization for empty car distribution. ORSA/TIMS National Conference, Detroit, MI., 1994.